

Controllability and stability of primary frequency control from thermostatic loads with delays

Charalampos ZIRAS¹ , Evangelos VRETTOS², Shi YOU¹



Abstract There is an increasing interest in exploiting the flexibility of loads to provide ancillary services to the grid. In this paper we study how response delays and lockout constraints affect the controllability of an aggregation of refrigerators offering primary frequency control (PFC). First we examine the effect of delays in PFC provision from an aggregation of refrigerators, using a two-area power system. We propose a framework to systematically address frequency measurement and response delays and we determine safe values for the total delays via simulations. We introduce a controllability index to evaluate PFC provision under lockout constraints of refrigerators compressors. We conduct extensive simulations to study the effects of measurement delay, ramping times, lockout durations and rotational inertia on the controllability of the aggregation and system stability. Finally, we discuss solutions for offering reliable PFC provision from thermostatically controlled loads under lockout constraints and we propose a supervisory control to enhance the robustness of their controllers.

Keywords Primary frequency control, Stability, Controllability, Delay, Thermostatically controlled loads

1 Introduction

The growing share of intermittent renewable energy sources (RES), especially wind power, is expected to increase the needs for ancillary services in power systems [1]. RES integration leads to gradual replacement of conventional power plants, which traditionally offer ancillary services to the grid. Thus, there is a growing interest in exploiting the flexibility of loads, commonly referred to as demand response (DR), to provide such services [2]. Furthermore, loads can respond faster than conventional generators, limiting the sudden frequency deviations caused by the reduced rotational system inertia due to a large share of resources which are connected to the grid through converters [3].

In this paper we focus on the provision of primary frequency control (PFC); PFC is a decentralized frequency control which acts quickly to contain the frequency deviation until secondary frequency control (SFC) restores the balance between generation and load [4]. Typically, full reserve must be provided within 30 seconds, proportionally to the frequency deviation. Thermostatically controlled loads (TCLs) are a common class of loads studied for offering ancillary services and more specifically PFC. A typical TCL is a load controlled by a thermostat, which keeps temperature within a deadband by switching on or off its heating or cooling element.

There is extensive research on how the consumption of TCLs can be manipulated so that their aggregate load can be controlled and offer ancillary services. However, many of these methods provide support in case of sudden and

CrossCheck date: 21 November 2016

Received: 30 August 2016 / Accepted: 21 November 2016 / Published online: 4 January 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

✉ Charalampos ZIRAS
chazi@elektro.dtu.dk

Evangelos VRETTOS
vrettos@eeh.ee.ethz.ch

Shi YOU
sy@elektro.dtu.dk

¹ Center for Electric Power and Energy, DTU, Lyngby, Denmark

² Power Systems Laboratory, ETH Zurich, Zurich, Switzerland



STATE GRID

STATE GRID ELECTRIC POWER RESEARCH INSTITUTE

large frequency deviations and do not meet the typical requirements of PFC. Some of the proposed methods are deterministic, and the devices' setpoints change based on the measured frequency in order to offer PFC [5, 6]. Although such methods provide a fast response to sudden frequency excursions, they tend to synchronize the loads and cause non-decreasing frequency oscillations [7].

A number of stochastic [8, 9] or hybrid solutions [10] (combining stochastic and deterministic elements) have been proposed. In [11] a multi-agent control approach is proposed, applying utility functions to offer primary and secondary reserves concurrently. In [12] a decentralized stochastic control method was presented, based on local frequency measurements, to offer PFC with an aggregation of refrigerators (REFs). The proposed controller accounts for the effect of increased power consumption due to the devices' startup dynamics and the compressors' lockout constraints. Such lockouts are used to avoid very frequent switching of the devices, which increases wear of the equipment and possibly reduces their lifetime.

Despite the plethora of control methods proposed, each with its own advantages and limitations, there is a gap in the literature concerning the dynamic response of loads. Specifically, most works in literature examine the ability of loads to track an exogenous frequency signal. Nevertheless, the integration of a significant share of DR-based PFC resources requires dynamic frequency studies, since the loads affect the evolution of frequency substantially.

In most works, the response time of the loads and the discretization step of the models are not thoroughly studied and dynamic simulations are not conducted. In [8] a time resolution of 10 s is used, which is too high to capture the frequency dynamics. In [13] loads are used to stabilize frequency in case of large disturbances, and it is shown that the response delay of loads plays a significant role in their ability to mitigate frequency deviations. Reference [14] indicates that delays over 0.5 s can lead to system instability, as the loads' controllers will react and change the aggregation's load out of phase with the electromechanical energy of the system generators.

In a realistic setup, loads would rely on frequency measurements on the low voltage grid in order to react to frequency deviations. In recent years power quality issues like harmonics and flicker have become more important [15], resulting in additional noise in the frequency signal. On top of that, the expansion of wireless communication networks results in higher interference with the power grid and higher noise levels. Therefore, due to low voltage power quality issues and the limited accuracy of local controllers (which must be cheap for wide-spread implementation), the measured frequency will likely be different than the actual system frequency. In [16] a study on the correlation of frequency measurements in the high and low

voltage grids was carried out. The observation was that there is a high correlation on the lower frequencies of the spectrum (until 0.4 Hz), but coherency drops for higher frequencies; this was attributed to the noise levels in the low voltage feeders. However, the authors concluded that frequency measured on the wall outlets can be used for participation in frequency control.

Even though demand response can be an alternative to conventional power plants offering PFC reserves, control design and the way loads react to frequency deviations must be carefully considered. It is outside the scope of the paper to represent the low voltage grid and the power quality phenomena related to it, whereas frequency is considered uniform in each area when conducting two-area power system simulations.

The contribution of this paper is fourfold. First, we propose a framework for the systematic consideration of frequency measurement and load response delay in DR-based PFC provision. Second, we introduce the concept of controllability index (*CI*), as a metric to evaluate PFC provision under lockout constraints, and we use it to design a robust controller. Third, we carry out extensive simulations to study the effects of measurement delay and rotational inertia on the controllability of an aggregation of REFs and system stability. Fourth, we propose a method based on multivariate regression to decrease the estimation errors of the number of locked on and off devices with lockout constraints, compared to [12].

The remainder of the paper is organized as follows. Section 2 provides the basic features of the decentralized control method which is presented in detail in [12]. In addition, Sect. 2 presents a new estimation method for the fraction of locked devices, which achieves better results compared with the estimation method of [12]. In Sect. 3 we propose a framework in order to systematically account for frequency measurement and response delays, and in Sect. 4 the simulation setup is described. In Sect. 5 the controllability index is introduced, and in Sect. 6 safe values for the response delays are determined via simulations. In Sect. 7 the results of extensive simulations are presented and solutions to increase the controllability margins are discussed. In Sect. 8 a discussion on how our findings and methods can be used by system operators is given, whereas Sect. 9 concludes.

2 Control and estimation methods

2.1 Control method

In this section an overview of the decentralized PFC method will be given, and some notation applied in the paper will be introduced; a detailed description of the control algorithm is given in [12]. The temperature of a

refrigerator i , operating with an on/off hysteresis between an upper and a lower thermostatic limit, is given by

$$\frac{d}{dt}T_i(t) = \alpha_i(T_{a,i} - T_i(t)) + w_i(t), \quad \text{if off} \quad (1)$$

$$\frac{d}{dt}T_i(t) = \alpha_i(T_{a,i} - T_i(t)) - \beta_i P_{n,i} + w_i(t), \quad \text{if on} \quad (2)$$

where $\alpha = 1/RC$; $\beta = \eta/C$; C is the thermal capacitance; R is the thermal resistance; T_a is the room temperature; η is the coefficient of performance; P_n is the nominal power. The noise term $w(t)$ represents the external disturbances, e.g. door openings. In discrete time, the temperature evolution can be described by

$$T_{t+1} = aT_t + bm_t + (1 - \alpha)T_a + w_t \quad (3)$$

where $a = e^{-\Delta t/(RC)}$; $b = (a - 1)\eta RP_n$; Δt is the discretization time step; $m_t \in \{0, 1\}$ is the on/off state.

The aggregate duty cycle is equal to the fraction of REFs that are at the on state and is expressed as a decimal $\in [0, 1]$. If the population is sufficiently large and the term w_t is neglected, then the nominal duty cycle D^n (and consequently the aggregate load) can be considered constant and is calculated based on the average parameters of the population. It is straightforward to modify the control algorithm so that it can be applied in the presence of external disturbances. However, in this work w_t is neglected as its inclusion does not substantially affect the studied phenomena. PFC can be offered by changing the population's desired duty cycle D_t^d , by stochastically switching a fraction of the loads according to

$$D_t^d = D^n + D^r \frac{\Delta f_t}{\Delta f_{\max}} \quad (4)$$

where Δf_t is the frequency deviation, Δf_{\max} is the frequency deviation for full PFC reserve activation, e.g. 0.2 Hz in continental Europe and D^r is the offered reserve capacity in terms of duty cycle. Such a stochastic approach cannot guarantee an accurate response to the frequency deviation if the population is small. However, due to the law of large numbers, an aggregation of a few thousands of loads (as shown in [12]) can provide PFC with sufficient accuracy. Indeed, the aggregate response can be very well described by the average values of the population parameters and the duty cycle can be approximated by (4). In the above formulation, a full PFC response in one time step is assumed. It is straightforward to extend the algorithm in order to achieve a response time equal to t_r seconds.

It is important to note that all calculations and estimations that will be presented are the same for each REF and refer to the aggregation as a whole, because it is assumed that the measured frequency is the same for all loads; therefore, the subscript i is not used. If the measured frequency is different for each load, then the calculations would be performed and applied for each REF individually.

Each REF measures the local Δf_t and calculates a switching probability ρ_t according to

$$\rho_t = \begin{cases} \frac{x_t}{1 - D_{t-1}^a - L_{\text{off},t-1}}, & \text{if } x_t \geq 0 \\ -\frac{x_t}{D_{t-1}^a - L_{\text{on},t-1}}, & \text{if } x_t < 0 \end{cases} \quad (5)$$

If $\rho_t \geq 0$, a number of devices have to switch on. If $\rho_t < 0$, a number of devices have to switch off. Each device draws a number $\in [0, 1]$ from a uniform distribution and changes state only if it is unlocked and in the opposite state, and the drawn number is smaller than ρ_t . x_t is the required change of the population's duty cycle between two time steps and D_{t-1}^a is the actual duty cycle at the previous time step. $L_{\text{on},t}$ and $L_{\text{off},t}$ are the estimated locked on and off load fractions respectively, with the random variables t_{on}^l and t_{off}^l representing the lockout durations of the loads. It must be noted that the desired and actual duty cycles of the population are not the same due to the inclusion of the startup dynamics, where the calculation of x_t takes into account past activations.

The control algorithm also modifies the thermostats' limits to improve performance in case of prolonged frequency deviations and contains a temperature feedback loop to achieve robustness in the case of biased frequency deviations. The thermostat limits are modified by incorporating an additional stochastic layer, to overcome practical limitations in the thermostat resolution [12]. These control actions do not substantially affect the behaviour of the aggregation as far as controllability is concerned, since they act on a much larger time scale. In the following section, we extend the method of [12] to improve the estimations of the locked fractions of the population estimation.

2.2 Estimation method

In [12], the estimation of $L_{\text{on},t}$ consisted of two parts; a constant steady-state part $L_{\text{on}}^{\text{st}}$ (attributed to the normal thermostat control and present even without reserves provision) and a transient part, $L_{\text{on},t}^{\text{tr}}$, attributed to the stochastic switching actions:

$$L_{\text{on},t} = L_{\text{on}}^{\text{st}} + L_{\text{on},t}^{\text{tr}} \quad (6)$$

A similar estimation was used for $L_{\text{off},t}$:

$$L_{\text{off},t} = L_{\text{off}}^{\text{st}} + L_{\text{off},t}^{\text{tr}} \quad (7)$$

The transient parts reflect the numbers of locked devices due to the stochastic switching actions of the controllers. $L_{\text{on},t}^{\text{tr}}$ and $L_{\text{off},t}^{\text{tr}}$ depend on past activations and can be estimated recursively assuming that the survival functions of the lockout durations are known; we denote them by $S_{\text{on}}(t_{\text{on}}^l)$ for the lock-on duration and $S_{\text{off}}(t_{\text{off}}^l)$ for the lock-



off duration. We further define N_{on} as the maximum lock-on duration and N_{off} as the maximum lock-off duration. Then, $L_{\text{on},t}^{\text{tr}}$ and $L_{\text{off},t}^{\text{tr}}$ can be computed as:

$$L_{\text{off},t}^{\text{tr}} = \sum_{k=0}^{t-1} c_k x_k S_{\text{on}}(t-k), \quad c_k = \begin{cases} 1 & \text{if } x_k \geq 0 \\ 0 & \text{if } x_k < 0 \end{cases} \quad (8)$$

$$L_{\text{off},t}^{\text{tr}} = \sum_{k=0}^{t-1} d_k x_k S_{\text{off}}(t-k), \quad d_k = \begin{cases} 0 & \text{if } x_k \geq 0 \\ 1 & \text{if } x_k < 0 \end{cases} \quad (9)$$

where $S_{\text{on}}(t-k) = 0$ for $t-k > N_{\text{on}}$, and $S_{\text{off}}(t-k) = 0$ for $t-k > N_{\text{off}}$.

While these equations provide a relatively accurate estimation, extensive simulations have shown that they result in offset errors. Even though the evolution of the locked fractions is captured quite accurately, an offset error is produced, which leads to a consistent overestimation of the locked on and off fractions. This offset error can be attributed to the fact that during reserve provision the loads that are locked due to the normal thermostatic control actions are fewer than $L_{\text{on}}^{\text{st}}$ and $L_{\text{off}}^{\text{st}}$.

A heuristic approach based on multivariate linear regression was adopted to improve the estimation. Simulations showed that very good results were obtained by using $L_{\text{on},t-1}$, $L_{\text{off},t-1}$, and D_t^a as predictors. A 24-hour simulation was used to calculate the parameters of the model; the regression terms $E_{\text{on},t}$ and $E_{\text{off},t}$ take the following form:

$$E_{\text{on},t} = g_1 + g_2 L_{\text{on},t-1} + g_3 L_{\text{off},t-1} + g_4 D_t^a \quad (10)$$

$$E_{\text{off},t} = q_1 + q_2 L_{\text{on},t-1} + q_3 L_{\text{off},t-1} + q_4 D_t^a \quad (11)$$

$E_{\text{on},t}$ is then added to the initial estimation (8) and the term $E_{\text{off},t}$ is added to (9), in order to achieve more accurate estimates. In Fig. 1 the estimation results are shown for a 5-hour simulation period with the population parameters

taken from [17]; $D^r = 0.178$, $\Delta t = 1$ s, whereas the frequency signal is an extract from the Swiss system (2011 data). The lock-on durations of the compressors follow a normal distribution with a mean value of 90 s and the lock-off durations a normal distribution with a mean value of 280 s. The actual locked fractions are shown with red colour and the estimation without the regression terms with a blue colour. The yellow lines show the improved estimation, achieved by the proposed method. It can be seen that the offset error is diminished and the estimated fractions of the locked loads are much closer to the actual values.

3 Frequency measurement and load response

In this section, we present a framework to enable the inclusion of frequency measurement and load response delay in reserve provision.

For such an application, a relatively cheap but highly reliable measurement device is required to measure local frequency. A simple method is to average frequency over a number of cycles, in a time period referred to as frequency measurement time (FMT) by using a zero-crossing method. There is a clear trade-off between speed and accuracy; the larger the time period used to measure frequency, the higher the accuracy. Typically a number of 5–10 cycles is required to maintain a good accuracy [18].

Apart from averaging the measured frequency, some additional processing of the raw data may be required to increase accuracy, namely a low-pass filter to reduce noise and interpolation between the zero-crossings [18]. After FMT has passed, a Processing and Response Time (PRT) is required where raw data is processed, a frequency value is calculated and the control algorithm runs in the device. All these various steps require some time to be executed and this time period is referred to as PRT.

The accuracy of frequency measurements presented in [18] achieved good precision, where measurements error presented a variance of 1.4 mHz. It must be noted that the ENTSO-E precision requirement is 10 mHz [19]. However, it is not clear how the measurement error can be represented, when taking it as a random variable superimposed on the actual frequency. A simple way to represent this stochasticity is via an uncorrelated Gaussian noise [11].

Figure 2 shows the sequence of measurement and response of the loads for random FMT and PRT values. In this case frequency samples are taken between $t = 100$ ms and $t = 300$ ms and for the next 300 ms the necessary data processing is done and the calculations of the control algorithm are performed, whereas the control actions are executed at $t = 600$ ms. The whole process is repeated every 300 ms. According to [20] FMT should not exceed 100 ms, which is the value used for all simulations.

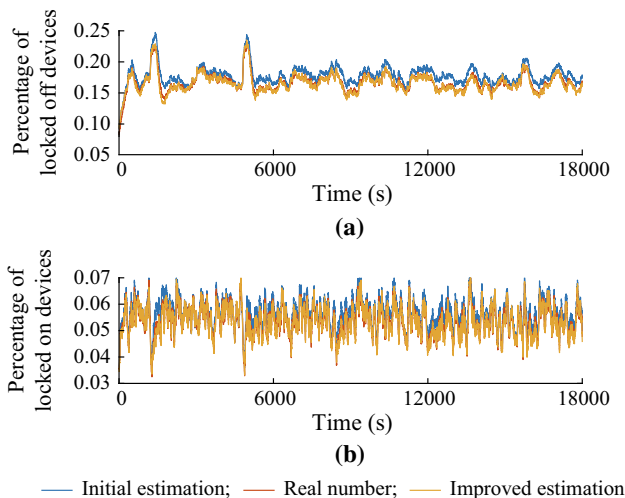


Fig. 1 Fractions of locked devices at the on and off states

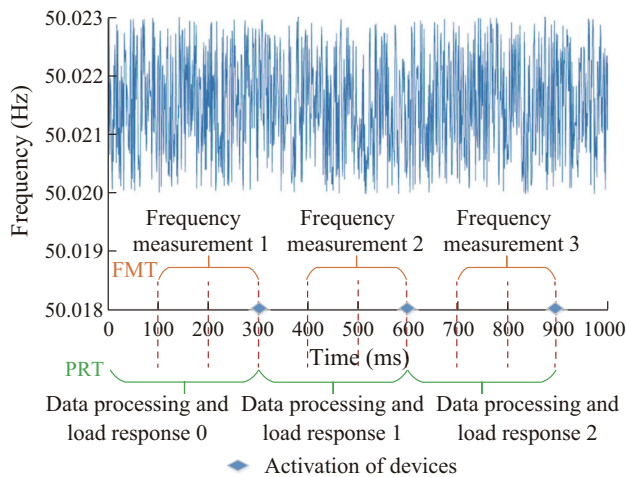


Fig. 2 Time sequence of frequency measurement and control response (FMT = 200 ms, PRT = 300 ms)

4 Simulation setup

We demonstrate the dynamic performance of an aggregation of REFs in PFC by using a linearized, two-area power system model [21], shown in Fig. 3. Area 1 is similar to the Swiss power system, whereas area 2 represents the rest of the interconnected European power system. The parameters of the two-area system are shown in Table 1.

The share of DR-based PFC of area 1 is denoted by α_1 and that of area 2 by α_2 . An aggregation of 50000 refrigerators was used in all simulations and its output was scaled up to match the PFC capacity while setting $D^f = 0.178$. For instance, an aggregation of 50000 refrigerators with a mean nominal power of 80 W and a mean duty cycle of $D^n = 0.25$ yields an average consumption of

Table 1 Two-area power system parameters

Parameter	Area 1	Area 2
Base load S_B (GW)	8	240
Damping D (MW/Hz)	120	3600
PFC reserves (MW)	71	2929
PFC droop $1/R$ (MW/Hz)	355	14645
Bias factor B (MW/Hz)	475	18245
AGC proportional gain C_p	0.17	0.17
AGC time constant T_N (s)	120	240
SFC reserves (MW)	400	14000

1 MW. The offered reserve capacity would be 0.7 MW and a scaling factor of 100 would be used to match the full PFC capacity of the Swiss system. It must be noted that this would require approximately 5 million refrigerators. Increasing D^f to 0.2 would reduce this number to 4.4 million, whereas similar devices with larger nominal power ratings (freezers or commercial beverage refrigerators) would further reduce the number of required devices.

The Automatic Generation Control (AGC) signal is calculated and dispatched every 4 s and the simulation time step (the system was discretized to allow Matlab simulations) is equal to 0.001 s. Fast (without reheat) and slow (with reheat) thermal power plants were used to provide SFC and the $1 - \alpha_1$, $1 - \alpha_2$ shares of conventional PFC. A rate limiter was used to represent the generators' ramping limitations. We conducted extensive simulations for varying system inertia values, lockout durations and loads' response times. The load mismatch for each area is shown in Fig. 4; the load mismatches were backwards calculated based on an actual 2011 Swiss frequency signal in such a way that conventional frequency control resulted in a

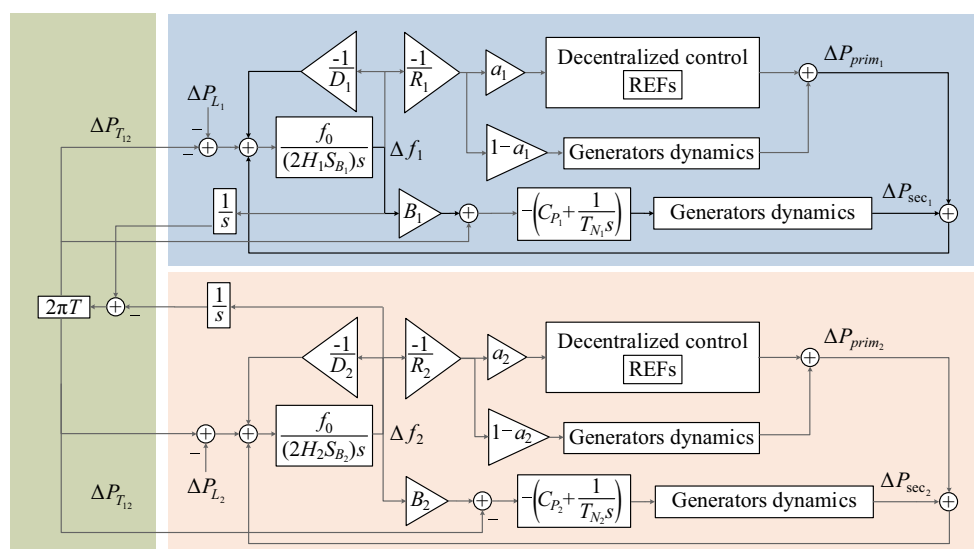


Fig. 3 Linearized two-area power system



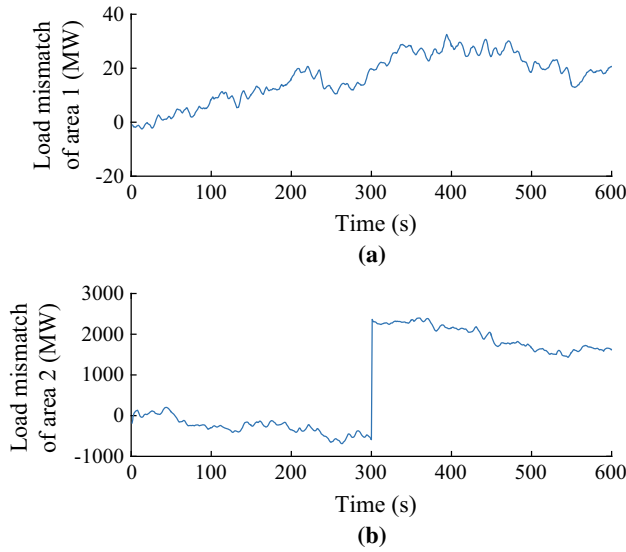


Fig. 4 Load disturbance for area 1 and area 2

response very similar to the measured frequency data. A sudden loss of generation in area 2, equal to 3000 MW, occurs at $t = 300$ s.

5 Aggregation controllability

If lockout constraints and startup dynamics are neglected, the loads can switch on and off arbitrarily often and are always able to offer frequency control accurately. In a more realistic setup, if such constraints are included, the load aggregation may become uncontrollable.

In general, loss of controllability can occur in both directions. Recalling the switching probability given by (5), reserves will be provided reliably as long as $\rho_t \in [0, 1]$. In other words, as long as there are enough on devices able to switch off ($-x_t \leq D_{t-1}^a - L_{on,t-1}$) and enough off devices able to switch on ($x_t \leq 1 - D_{t-1}^a - L_{off,t-1}$). For an average duty cycle of approximately 0.25 it is more likely to lose controllability in the positive PFC direction, i.e., while reducing aggregate power. The controllability index CI is defined by considering both directions as

$$CI_t = \min(D_{t-1}^a - L_{on,t-1}, 1 - D_{t-1}^a - L_{off,t-1}) \quad (12)$$

When CI_t becomes negative, then controllability is lost, as the population is not able to follow the frequency deviation. In the control algorithm presented in [12] loss of controllability was not considered because simulations with an exogenous frequency signal were performed. When DR-based reserves comprise the majority of fast frequency reserves, then dynamic simulations using a power system model are required, because frequency is substantially affected by the behaviour of these sources.

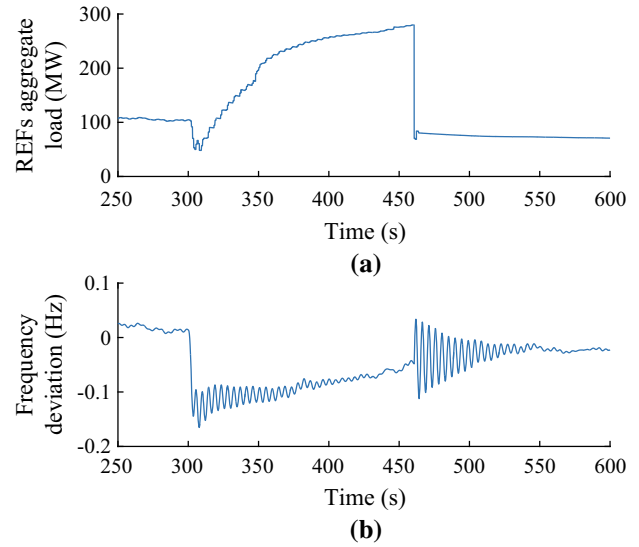


Fig. 5 Area 1 REFs aggregate load and frequency deviation for disturbances shown in Fig. 4

Following a large disturbance, and if the loads have a large response delay, then the resulting oscillations may lock large fractions of the population and render the aggregation uncontrollable. If this is not accounted for in the control of the devices, then they might exhibit an extreme behaviour when CI_t is very close to zero or negative, as ρ_t will take very large or negative values. Since ρ_t corresponds to a random number in the range $[0, 1]$, values outside this range will lead to an uncontrollable operation of the loads. We refer to this operation mode as “no CI limit”, as no limit is imposed on CI_t .

Negative values of CI lead to an extreme behaviour of the algorithm as shown in Fig. 5, for a sudden load disturbance; PRT was set to 0.9 s and FMT equal to 0.1 s, system inertia constants $H_1 = H_2 = 6$ s and $\alpha_1 = \alpha_2 = 1$, so that no power plants offer PFC. Loads fully respond to a frequency deviation at the next PRT step. Negative CI_t disassociates the switching probability and control actions from x_t and while $L_{on,t}$, $L_{off,t}$, and D_t^a are updated, their values do not correspond to the actual ones. A way to solve this problem is to impose a CI limit, which prohibits negative CI values and allows the aggregation to provide only the reserve capacity it is able to offer. If for a negative x_t it holds $-x_t \geq D_{t-1}^a - L_{on,t-1}$, then x_t will take a value such that $CI_t = 0$, in other words $x_t = -(D_{t-1}^a - L_{on,t-1})$. Similarly, if for a positive x_t it holds $x_t \geq 1 - D_{t-1}^a - L_{off,t-1}$, then $x_t = 1 - D_{t-1}^a - L_{off,t-1}$.

The CI limit significantly improves the behaviour of the aggregation, as shown in Fig. 6. Even though the large fraction of locked devices does not allow the aggregation to follow the frequency signal accurately, the switching probability ρ_t always corresponds to the associated x_t and

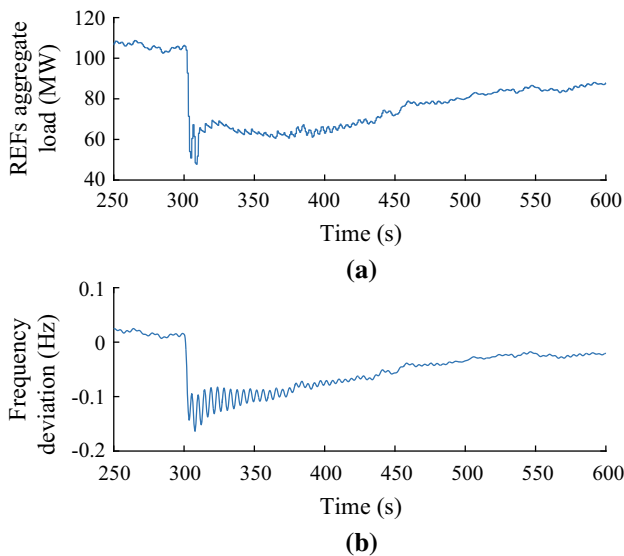


Fig. 6 Area 1 REFs aggregate load and frequency deviation for disturbances shown in Fig. 4, imposing a *CI* limit

the algorithm is stable. However, due to the critical nature of PFC reserves to the stability and reliability of the power system, loss of PFC controllability should not be tolerated under any circumstances.

6 Determining safe regions of *PRT*

6.1 Immediate loads response

We first examine the behaviour of the system for typical inertia values ($H_1 = 6$ s, $H_2 = 6$ s), fast generators providing PFC and SFC and varying *PRT* values, whereas *FMT* was set equal to 0.1 s for both areas. We denote with PRT_1 and PRT_2 the processing and response times of areas 1 and 2 respectively.

In this section the system behaviour is studied, when the lockout constraints and the startup dynamics are neglected. This is the most favourable case for the loads, because no constraints are imposed on their performance. If a certain *PRT* value induces an unstable system operation under these considerations, then this value is considered unacceptable. The ramping of the loads is set equal to one *PRT* step, which means that loads will fully respond to a frequency deviation after this time period has passed.

In Fig. 7, the frequency deviation is shown for 3 cases; the frequency nadir in the case of conventional PFC is 49.7 Hz, whereas DR-based PFC with a *PRT* equal to 0.2 s reduces the under-frequency value to 49.85 Hz due to the loads' faster response. Generators are not able to respond so quickly due to their ramping limitations, resulting in a considerably larger under-frequency. DR-based PFC is also

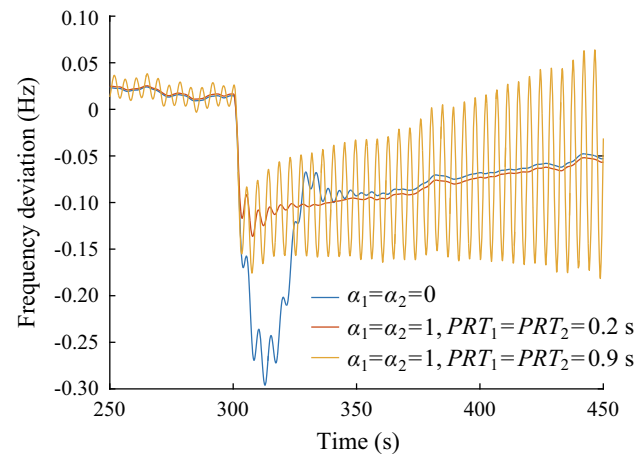


Fig. 7 System frequency of area 1 with $H_1 = H_2 = 6$ s and immediate loads response

able to damp the oscillation of the tieline power flow between the two areas much faster, within 25 s.

When *PRT* is set equal to 0.9 s the tieline oscillation is reinforced, gradually increasing in amplitude and activating more PFC reserves over time, which eventually leads the system to an unstable operation (*PRT* values smaller or equal to 0.8 s did not lead to instability). The period of the oscillation is approximately 0.2 s, corresponding to the frequency of the inter-area mode. Simulations showed that a relatively large *PRT* combined with a full reserve activation in one time step can lead to system instability.

Smaller system inertia values require even smaller *PRT*s for a safe system operation due to more volatile frequencies and tieline power flows. For the studied power system, *PRT* values larger than 0.5 s led to instability for inertia values equal to 3 s. The improvement of DR-based PFC over conventional generators in the frequency profile is even more significant as shown in Fig. 8. The fast response of the loads is able to substantially contain the frequency deviation within the first seconds, largely cancelling the negative effect of the small inertia.

6.2 Introduction of ramping profile

As shown in the previous examples, the combination of a relatively large *PRT* and an immediate response may result in an unstable operation. Introducing a ramping profile of a few seconds leads to the damping of the power oscillations caused by the sudden disturbance.

In Fig. 9, the frequency deviation of area 1 is shown, when a ramping time $t_r = 3$ s is applied. Ramping time can be seen as a controller's gain in PFC; the larger the t_r the smaller the gain of the controller. The system performance is considerably different, compared to that of a ramping time equal to one *PRT* (shown in Fig. 9). The slower

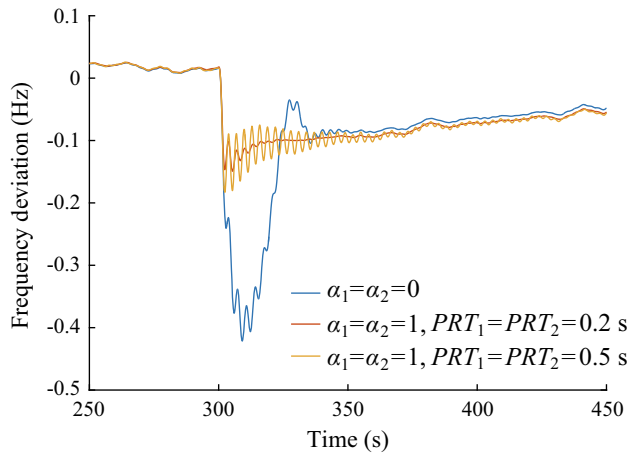


Fig. 8 System frequency of area 1 with $H_1 = H_2 = 3$ s and immediate loads response

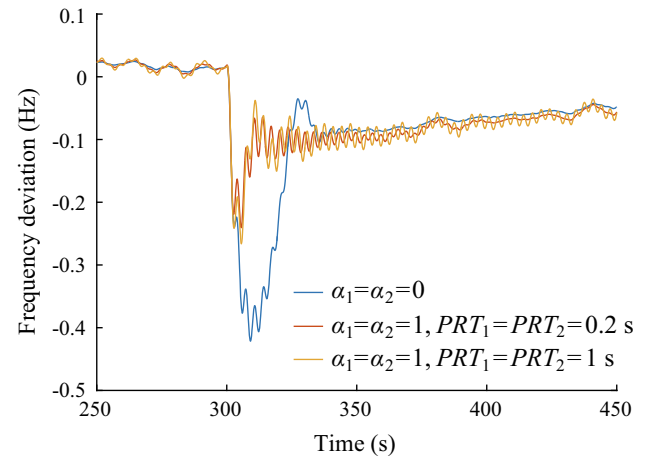


Fig. 10 System frequency of area 1 with $H_1 = H_2 = 3$ s with a ramping time equal to 3 s

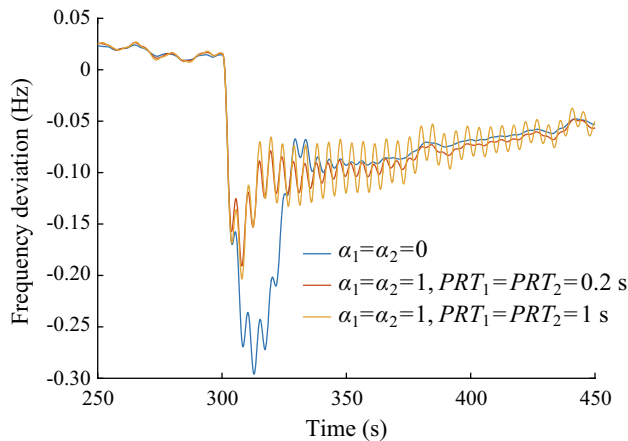


Fig. 9 System frequency of area 1 with $H_1 = H_2 = 6$ s with a ramping time equal to 3 s

response leads to a larger under-frequency for both response times. If $PRT = 0.2$ s, the oscillation is damped over a significantly larger period of 80 s, compared to 25 s when loads responded within one PRT . However, the larger t_r does not introduce a negative damping and the system remains stable even when $PRT = 1$ s.

In Fig. 10, the frequency deviation of area 1 is shown when a smaller system inertia value of 3 s is used. Although the frequency drop is larger, the ramping behaviour of the loads results in stable operation for larger PRT s and small inertia values.

These examples indicate that a very fast response reduces the maximum value of frequency deviation. However, if this very fast response is combined with a critically high delay, then the system becomes unstable. A larger t_r does not introduce negative damping for the same delays and allows the system to be stable for relatively large PRT values. The disadvantage of this method is that

the maximum frequency deviations are larger and the resulting oscillations have increased settling times, compared to the case where an immediate full response is allowed. The best behaviour from the power system's perspective seems to occur when a very fast response is combined with a minimal delay. In that case frequency deviations are significantly limited and oscillations are damped very quickly.

7 Simulation results

7.1 DR performance under lockout constraints and startup dynamics

We now consider the system operation when the startup dynamics and the lockout constraints are not ignored and the control algorithm takes them into account. We examined 49 cases, for varying values of $PRT_1 \in [0.2, 0.8]$ and $PRT_2 \in [0.2, 0.8]$ with a step of 0.1 s to determine when load controllability is lost; simulations showed that values above 0.8 s are considered unacceptable for safe system operation and were therefore not studied.

The first set of simulations was conducted with $H_1 = H_2 = 6$ s and $\alpha_1 = \alpha_2 = 1$ and full response within one PRT step. As seen in Fig. 11, REFs lose controllability for $PRT \geq 0.3$ s. The zero values of CI_t indicate that the CI limit control is activated, limiting the offered reserve to values smaller than the requested. Even in the case of $PRT = 0.2$ s, the controllability margin is very low and an additional disturbance or a smaller system inertia value would lead to the activation of the CI limit.

Figure 12 presents the results with $t_r = 3$ s. In this case controllability margins are larger and controllability is retained even for larger PRT values. This is attributed to

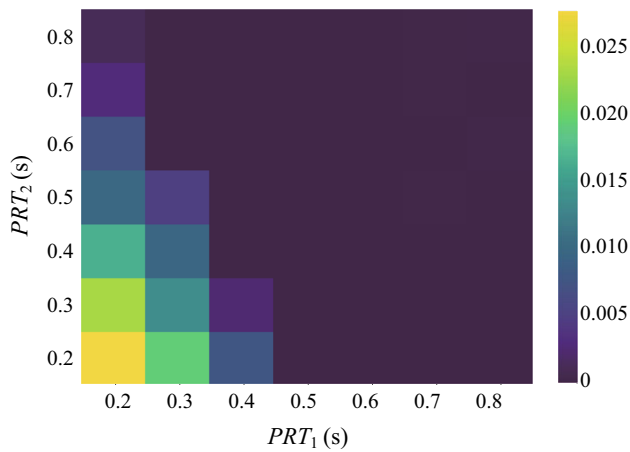


Fig. 11 Minimum values of CI_t for area 1 with $H_1 = H_2 = 6$ s for full PFC activation within one time step

the slower response of the loads, which leads to smaller fractions of locked devices and larger CI_t values. Further simulations were conducted using slower generators and smaller inertia values. The dynamics of the generators did not significantly affect the behaviour of the system, whereas smaller inertia values result in much worse CI values and system performance. From the conducted simulations we can conclude that ON/OFF resources limited by lockout constraints are susceptible to loss of controllability in case of large load disturbances when D^f is high. There is a number of potential solutions to increase the controllability margins of the aggregation, which are presented in the following subsection.

7.2 Controllability improvement

A smaller D^f value would activate smaller fractions of the population and would lead to fewer locked devices, increasing the controllability margins. However, D^f is a

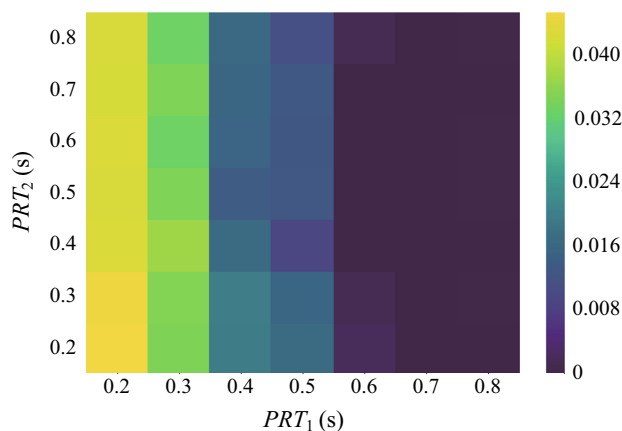


Fig. 12 Minimum values of CI_t for area 1 with $H_1 = H_2 = 6$ s for a ramping time of 3 s

value very critical to the economic viability of the PFC provision of such loads. A decrease of D^f from 0.178 to 0.15 would require 18% more devices, with the safe system operation still not guaranteed.

Another possibility is to decrease the lockout durations of the compressors, especially at the on state. This would lead to smaller locked on fractions and larger controllability margins. However, significantly reducing these durations might cause wear to the devices and potentially decrease their lifetime. Next, simulations are carried out with reduced lockout durations and the concept of a controllability supervisory control is introduced and tested via simulations.

7.2.1 Reduced lockout durations

Lock on and lock off durations were decreased in order to examine the effect of the reduction on the population's controllability; the mean value of the lock on durations is now set to 45 s and the mean value of the lock off durations is equal to 140 s. In Fig. 13 the results for an immediate response are shown.

As expected, CI_t increased as the smaller lockout durations reduce the number of locked devices. Loss of controllability occurred in larger PRT values but again the margins were very low, whereas a reduction of system inertia results in loss of controllability in all cases except for $PRT = 0.2$ s. A slower response significantly improves CI_t , as shown in Fig. 14.

However, as already mentioned, an immediate loads response results in smaller frequency deviations and shorter settling times of the tieline oscillations. But even if a ramping time of 3 s is used, the robustness of the control is not ensured. Small system inertia values, inter-area oscillations or sudden load disturbances may result in loss of

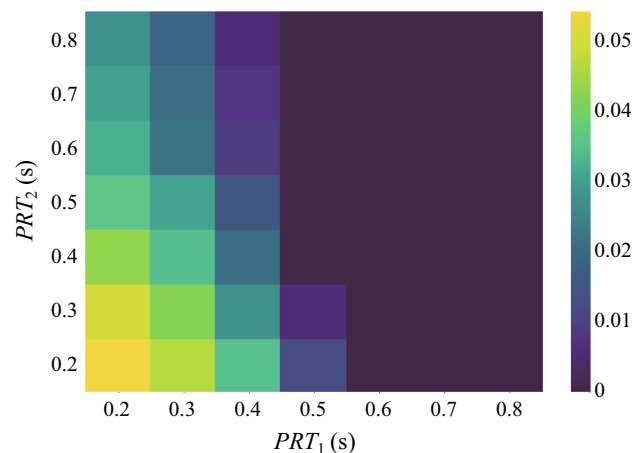


Fig. 13 Minimum values of CI_t for area 1 with $H_1 = H_2 = 6$ s for full PFC activation within one time step and reduced lockout durations

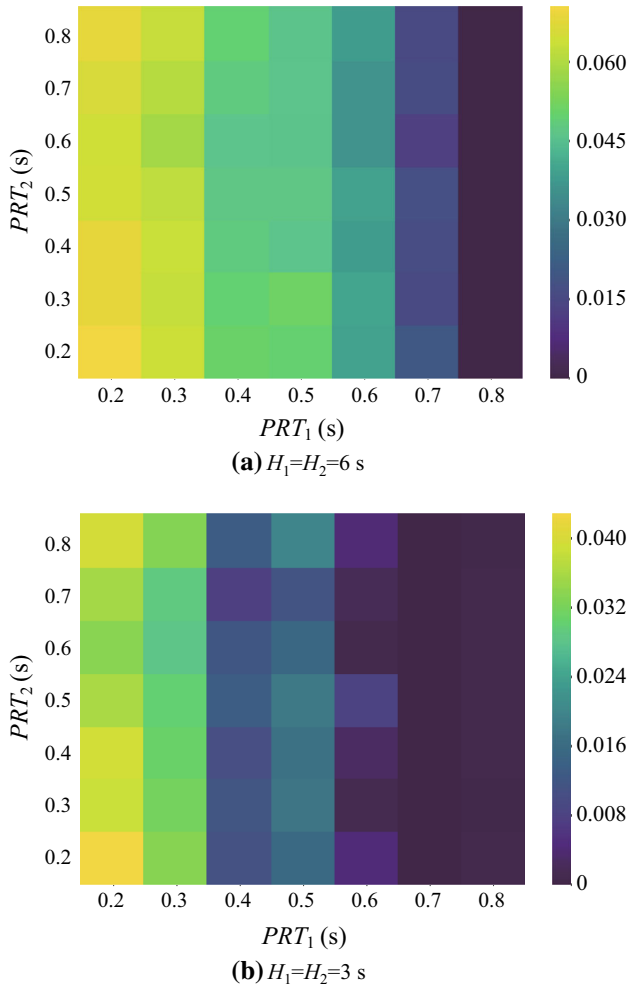


Fig. 14 Minimum values of CI_t for area 1 for a ramping time of 3 s and reduced lockout durations

controllability. It is necessary to introduce a controllability supervisory control (CSC), which will monitor CI_t and will unlock a fraction of the locked devices to ensure that the required reserves are always provided.

7.2.2 Controllability supervisory control

The purpose of CSC is to unlock the necessary number of loads so that the population never loses controllability. Generally, this is undesired because it will cause wear to the devices. However, CSC is essential in order to provide PFC reliably, and it would be very rarely activated, only in case of very large disturbances. If for a negative x_t it holds $-x_t \geq D_{t-1}^a - L_{on,t-1}$, then $L_{on,t}$ must decrease by $D_{t-1}^a + x_t$. To calculate the percentage of locked on devices ξ_t that must become unlocked, this number must be divided by the available loads to unlock, which are equal to $L_{on,t-1}$. Similar arguments hold for a positive x_t and locked off devices. The unlocking probability ξ_t is thus given by

$$\xi_t = \begin{cases} \frac{L_{on,t-1} - D_{t-1}^a - x_t}{L_{on,t-1}}, & \text{if } -x_t \geq D_{t-1}^a - L_{on,t-1} \\ \frac{1 - D_{t-1}^a - L_{off,t-1} + x_t}{L_{off,t-1}}, & \text{if } x_t \geq 1 - D_{t-1}^a - L_{off,t-1} \end{cases} \quad (13)$$

CSC can take various forms. One solution is, for a negative x_t , the loads to generate a random number $\in [0, 1]$ and unlock if the drawn number is lower than ξ_t and if the load is locked on. This will cause a uniform unlocking of the devices, irrespective of the time that has passed since each device locked. Another approach would be to deterministically unlock a ξ_t fraction of the locked on loads, based on the elapsed time since they locked. This, however, requires a uniform distribution for the elapsed periods since the lockout, which does not necessarily hold, and could be problematic if applied several times over a short time period, due to the changes caused on the distributions. In both cases, a similar logic can be applied for the locked off loads.

The former solution was adopted, but instead of unlocking a certain fraction of the locked devices given by ξ_t , a CI_t threshold equal to 0.05 was used, where all locked on or locked off loads are unlocked (depending on which part is required to unlock). This was done to avoid states with marginal controllability, which are undesired due to the stochasticity of the control. To test the CSC, simulations were conducted for $H_1 = H_2 = 6$ s, $PRT_1 = PRT_2 = 0.6$ s, with reduced lockout durations and an immediate response. As previously shown in Fig. 13, without CSC these parameters resulted in loss of controllability. Figure 15 compares the behaviour of the aggregation with and without CSC.

Without CSC, at approximately 320 s controllability of the aggregation is lost and CI_t is zero, as the loads offer as much reserve as possible, without being able to follow the frequency deviation signal. On the other hand, CSC lifts the lockout constraints of the locked on devices when $CI_t = 0.05$, increasing the controllability margin. The aggregation is thus able to follow the frequency deviation, as shown in Fig. 15 a, which in turn results in faster-decaying power oscillations. Moreover, due to the controllability margin, the aggregation is able to respond to additional frequency deviations, which is not the case when CSC is not applied.

8 Discussion

The method we have presented can be used by system operators to define requirements for the safe integration of TCLs in PFC. To do so, this would require more accurate power system dynamic models (with multiple areas and higher modelling detail) and a careful consideration of

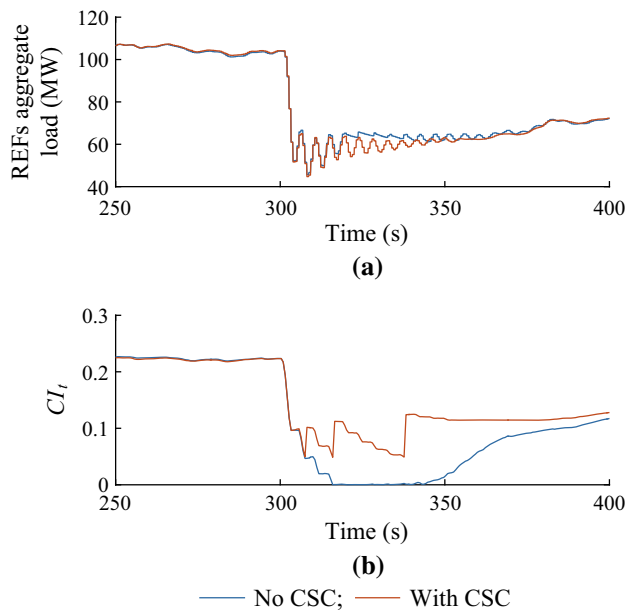


Fig. 15 Aggregate REFs load and CI_t with and without CSC

rotational inertia values. We have demonstrated that lower inertia values, which occur during periods of high penetration of RES, further reduce the controllability margins and require faster response times to avoid unstable operation. Such simulations can provide maximum response times for the measurement of frequency and activation or deactivation of devices. Once these response times have been set as requirements for DR PFC providers, they can be tested in pre-qualification tests, before frequency-responsive loads can be safely integrated in frequency reserves provision.

We have further demonstrated the possibility of loss of controllability, under certain circumstances such as long lockout durations, low inertia values, very low loads ramping times, long delays and high offered reserve capacity. Some of these parameters may be determined by external factors, such as low system inertia values, REFs lockout times set by the manufacturers, or FMT set by the measurements devices. Moreover, a large reserve capacity (compared to the loads' baseline) is financially more attractive for the aggregation. We proposed a CSC which ensures that controllability over the population is retained, by monitoring the controllability margins and increasing them when needed, by unlocking parts of the population. This method can overcome the practical limitations imposed by the above mentioned factors and ensure the controller's robustness.

Finally, we have showed that if a ramping time of a few seconds is introduced, the allowed delays can be longer, but from a power system's perspective an

immediate response results in better frequency profiles (smaller deviations and settling times for the power oscillations). Therefore, there is a clear trade-off between the loads' response delay and their ramping rate. If the response delay is relatively large due to practical limitations (for instance due to a long FMT), then the ramping rate must decrease, to avoid large power oscillations and unsafe system operation. This trade-off can be incorporated in the PFC scheme by defining certain allowed regions for combinations of response delays and ramping rates. The operator would enforce lower ramping rates to aggregations which exhibit larger response delays and would allow faster ramping rates when loads can respond with smaller delays. Since a higher ramping rate improves the system's performance (provided that the delay is sufficiently small), a performance payment should be followed, where these resources would be remunerated with higher prices, to ensure a fair compensation to the providers.

9 Conclusion

In this paper, we investigated the effects of response delays and lockout constraints on the controllability of an aggregation of REFs offering PFC. We used a two-area power system model to provide insight on how these important parameters affect PFC provision. We showed that loads which fully respond to a frequency deviation must do so with maximum delays in the order of a few hundreds of ms, otherwise they might induce an unstable operation.

If lockout constraints are present, then the controllability of the aggregation might be lost after a sudden disturbance. The conditions where controllability is lost depend on the system inertia values, the loads' ramping time, the delays, and the devices' lockout durations. Simulations showed that even for small delays controllability margins might be very low. Various solutions were discussed and it was concluded that a supervisory control is the most reliable way to enhance robustness.

Such a control, which monitors the controllability margins and unlocks the devices in case these margins become critically low, allows the aggregation to offer reserves in a robust manner. Moreover, considering that very large disturbances are rare and such a control would be activated only on these rare occasions, it is possible to maximize the offered reserve capacity. Indeed, if controllability is guaranteed via the CSC, reserve capacity could be set almost equal to the mean duty cycle of the loads, which would be a considerable improvement of their business model.

Acknowledgements C. ZIRAS and S. YOU acknowledge the financial support of the EUDP funded project Ecogrid 2.0. E. VRETTOS acknowledges the financial support of Nano-Tera.ch for the HeatReserves project and of the Swiss Federal Office of Energy and Swisselectric Research for the SmartGrid-Polysun: Design Tool for Local Load Management project.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- [1] Makarov YV, Loutan C, Ma J et al (2009) Operational impacts of wind generation on California power systems. *IEEE Trans Power Syst* 24(2):1039–1050
- [2] Callaway DS, Hiskens IA (2011) Achieving controllability of electric loads. *Proc IEEE* 99(1):184–199
- [3] Zhang S, Mishra Y, Shahidehpour M (2016) Fuzzy-logic based frequency controller for wind farms augmented with energy storage systems. *IEEE Trans Power Syst* 31(2):1595–1603
- [4] Rebours YG, Kirschen DS, Trotignon M et al (2007) A survey of frequency and voltage control ancillary services part I: technical features. *IEEE Trans Power Syst* 22(1):350–357
- [5] Short JA, Infield DG, Freris LL (2007) Stabilization of grid frequency through dynamic demand control. *IEEE Trans Power Syst* 22(3):1284–1293
- [6] Xu Z, Østergaard J, Tøgeby M (2011) Demand as frequency controlled reserve. *IEEE Trans Power Syst* 26(3):1062–1071
- [7] Angeli D, Kountouriotis PA (2012) A stochastic approach to dynamic-demand refrigerator control. *IEEE Trans Cont Syst Technol* 20(3):581–592
- [8] Borsche T, Markovic U, Andersson G (2016) A new algorithm for primary frequency control with cooling appliances. *Comput Sci Res Dev* 31(1–2):89–95
- [9] Moghadam MRV, Ma RTB, Zhang R (2014) Distributed frequency control in smart grids via randomized demand response. *IEEE Trans Smart Grid* 5(6):2798–2809
- [10] Tindemans SH, Trovato V, Strbac G (2015) Decentralized control of thermostatic loads for flexible demand response. *IEEE Trans Control Syst Technol* 23(5):1685–1700
- [11] Weckx S, Hulst RD, Driesen J (2015) Primary and secondary frequency support by a multi-agent demand control system. *IEEE Trans Power Syst* 30(3):1394–1404
- [12] Vrettos E, Ziras C, Andersson G (2016) Fast and reliable primary frequency reserves from refrigerators with decentralized stochastic control. *IEEE Trans Power Syst*. doi:[10.1109/TPWRS.2016.2630601](https://doi.org/10.1109/TPWRS.2016.2630601)
- [13] Samarakoon K, Ekanayake J, Jenkins N (2012) Investigation of domestic load control to provide primary frequency response using smart meters. *IEEE Trans Smart Grid* 3(1):282–292
- [14] Donnelly M, Trudnowski D, Mattix S et al (2012) Autonomous demand response for primary frequency regulation. Pacific Northwest National Laboratory, Washington, Technical Report PNNL-21152
- [15] Liu YJ, Chang TP, Chen HW et al (2014) Power quality measurements of low-voltage distribution system with smart electric vehicle charging infrastructures, pp 631–635
- [16] Donnelly M, Harvey D, Munson R et al (2010) Frequency and stability control using decentralized intelligent loads: benefits and pitfalls, pp 1–6
- [17] Ziras C, Vrettos E, Andersson G (2015) Primary frequency control with refrigerators under startup dynamics and lockout constraints. In: 2015 IEEE power energy society general meeting, pp 1–5
- [18] Douglass PJ, Garcia-Valle R, Nyeng P et al (2011) Demand as frequency controlled reserve: implementation and practical demonstration. In: 2011 2nd IEEE PES international conference and exhibition innovative smart grid technologies (ISGT Europe), pp 1–7
- [19] Schlipf AD, Scherer M, Haller M (2011) Test for primary control capability. <https://www.swissgrid.ch/>
- [20] Energinet.dk (2011) Batteries as supplier of frequency-controlled reserves. Technical Report. 83861/11 v1
- [21] Kundur P, Balu NJ, Lauby MG (1994) Power system stability and control. McGraw-hill, New York

Charalampos ZIRAS received the Dipl.-Ing degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 2009. He then worked as a telecommunications engineer on benchmarking and optimization of mobile networks. In 2015 he received the M.Sc. degree in Energy Science and Technology from ETH Zurich (Swiss Federal Institute of Technology), Switzerland. In 2016 he joined the Center for Electric Power and Energy of DTU (Technical University of Denmark), where he is currently pursuing his Ph.D. His research interests include aggregation, optimization, and control of diversified portfolios of distributed energy resources, concurrent provision of power system services, and interaction with the distribution grid.

Evangelos VRETTOS received the Dipl.-Ing degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 2010. After his graduation, he worked as a research assistant at the Electric Machines and Power Electronics Laboratory of NTUA. In 2016 he received the Ph.D. degree from ETH Zurich (Swiss Federal Institute of Technology), Switzerland, where he worked at the Power Systems Laboratory. His research interests include control theory and optimization methods to enable efficient demand response programs for integration of renewable energy sources in the power system. He is also interested in energy storage technologies and energy-efficient control of buildings.

Shi YOU received his M.Sc. and Ph.D. in Electrical Engineering from Chalmers Institute of Technology, Sweden in 2007 and DTU (Technical University of Denmark) in 2010 respectively. Currently, he is a Research Scientist at the Center for Electric Power and Energy, Department of Electrical Engineering at DTU. His main research interests include market-based control, management and integration of distributed energy resources, planning, operation and management of active distribution networks and integrated energy systems.